Social Fragmentation at Multiple Scales

Leila Hedayatifar, **Yaneer Bar-Yam**, **Alfredo J. Morales**
New England Complex Systems Institute, Cambridge, MA, 02139

September 21, 2018

Abstract

Despite global connectivity, societies seem to be increasingly polarized and fragmented. This phenomenon is rooted in the underlying complex structure and dynamics of social systems. Far from homogeneously mixing or adopting conforming views, individuals self-organize into groups at multiple scales, ranging from friends and families up to cities and cultures. In this paper, we study the fragmented structure of the American society using mobility and communication networks obtained from social media data. We find self-organized patches with clear geographical borders that are consistent between physical and virtual spaces. The patches have multiscale internal structure from parts of a city to the entire nation. Our observations are consistent with the emergence of social groups whose separated association and communication reinforce distinct identities. Multiscale identities indicate progressively stronger association at smaller scales, but allow for self-association with larger groups at larger scales. Those who are "foreigners" at finer scales, may be part of the same group at larger scales. Understanding the emergence of fragmentation in hyper-connected social systems is imperative in the age of the Internet and globalization.

The increasing polarization of societies is becoming apparent around the world. Despite access to global communication [1], people seem to be splitting into groups that mostly listen to their own members [2, 3, 4]. It is crucial to understand the properties and underlying mechanisms of social fragmentation, given that it affects the way information flows among individuals [5] and consequently their emergent behaviors [6, 7, 8] including political or physical conflict [9, 10, 11, 12].

Individual choices of association play a major role in the emergence of social fragmentation [13, 8]. These choices can be influenced by ideologies [6, 14, 15, 16], occupations [17, 8] or consumer habits [18]. However, the structural properties of collective association and its relationship to the social space remains unclear.

The social space is a place where people frequently meet, interact with each other, create communities, and share values and behaviors [19]. While group cohesion is strongly influenced by internal communication, weaker external ties are responsible for integration at larger scales, providing individuals with information and resources beyond the borders of their own community [20, 21]. Previous studies have shown that the structure of strong and weak ties affects the behavior of social systems, including the spread of innovation [22], business and culture [23], as well as the development of regional and national events [21].

The recent availability of large-scale data sets obtained from mobile phones and social media has considerably improved our ability to study social systems [24, 25, 26, 27, 28, 29, 30, 31, 32, 33]. They enable direct observation of social interactions and collective behaviors with unprecedented detail. In this work we analyze human mobility and communication networks in order to describe and understand the multi-scale structure of social fragmentation in the US. We also propose a model to explain and frame the patterns we observe in the data based on natural human dynamics and network growth.

We use geo-located Twitter data to generate geographical networks based on where people travel or communicate. Nodes represent a lattice of 0.1° latitude $\times$ 0.1° longitude cells overlaid on a map of the U.S. Each cell is approximately 10 km wide. Network edges reflect two types of data: mobility and communication. In the mobility network, edges are created when a user u tweets consecutively from two locations i and j. In the communication network, edges are created when a user u at location i mentions another user v that has most recently tweeted at location j. The weight of an edge represents the number of people who either travel or communicate between i and j. This network aggregates the heterogeneities of human activities in a large-scale representation of social collective behaviors [34].

In Figure 1, we show the spatial properties of the mobility network in terms of degree centrality (Fig. 1-a) and two levels of modular structure (Fig. 1-b and Fig. 1-c). The degree centrality shows the density of user movements at each geographical point. The activity is concentrated in large cities (red in Fig. 1-a) and decreases toward suburban and rural areas (green, blue and gray). In areas of the country with high population density, cities merge into large regions of high activity (e.g., the East Coast corridor). In other areas, roads are also visible, as people tweet when they travel between cities. Highways in rural areas with higher traffic appear in green and less traveled roads are blue.

The spatial fragmentation of social systems arises when people travel and choose which boundaries not to cross either directly or incidentally. We study fragmentation by means of the modular structure of the network (Figure 1-b), obtained from modularity optimization [35]. Our results suggest that the US is fragmented into 20 large regions whose boundaries often follow state boundaries but may in particular cases be parts of one state, a single state or the combination of multiple states. For example, New England is mostly a single area, Georgia and Alabama are combined, as are Arizona and New Mexico, and North and South Carolina. Oregon and Washington as well as most of Idaho and Montana are combined in a large Northwest area. Another combined area is formed from North and South Dakota, Nebraska, Minnesota, Iowa and Northwest Wisconsin. These two combined state areas are larger than the area of Texas which itself is a single region, though it looses parts of West Texas to the Arizona and New Mexico region. New York City is combined with much of New Jersey and eastern Pennsylvania, while most of northern New York state is a separate region connected to parts of northern Pennsylvania. Oklahoma is combined with Kansas and Missouri. Louisiana, Mississippi, Arkansas and parts of south west Tennessee are a single area. Southern California is separate from Northern California.

At a finer scale of subdivision, these large regions are subdivided into patches that typically include individual cities and their surrounding areas (see Figure 1-c). There are 206 such communities that we obtain by applying the same modularity optimization algorithm to each larger

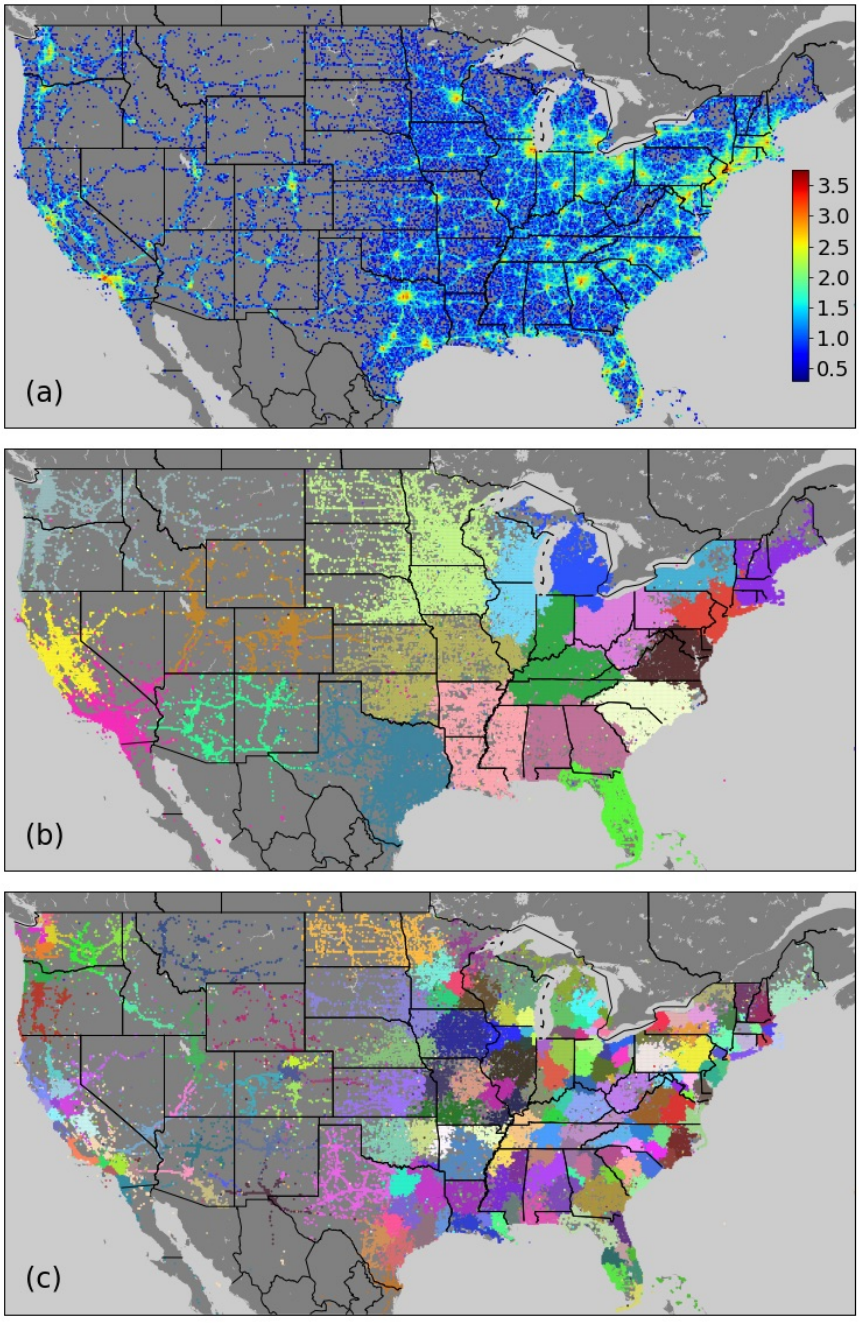


Figure 1: Structure and fragmentation patterns of the network associated with human mobility. (a) Spatial degree centrality of the mobility network. Colors indicate the amount of people traveling at each location, measured by the logarithm of the degree centrality of each node (scale inset). The mobility network was used to generate communities using modularity optimization, with distinct colors indicating (b) 20 patches that can be visually associated to states or regions and (c) 206 smaller sub-communities within the communities of panel (b) that can be visually associated to urban centers.

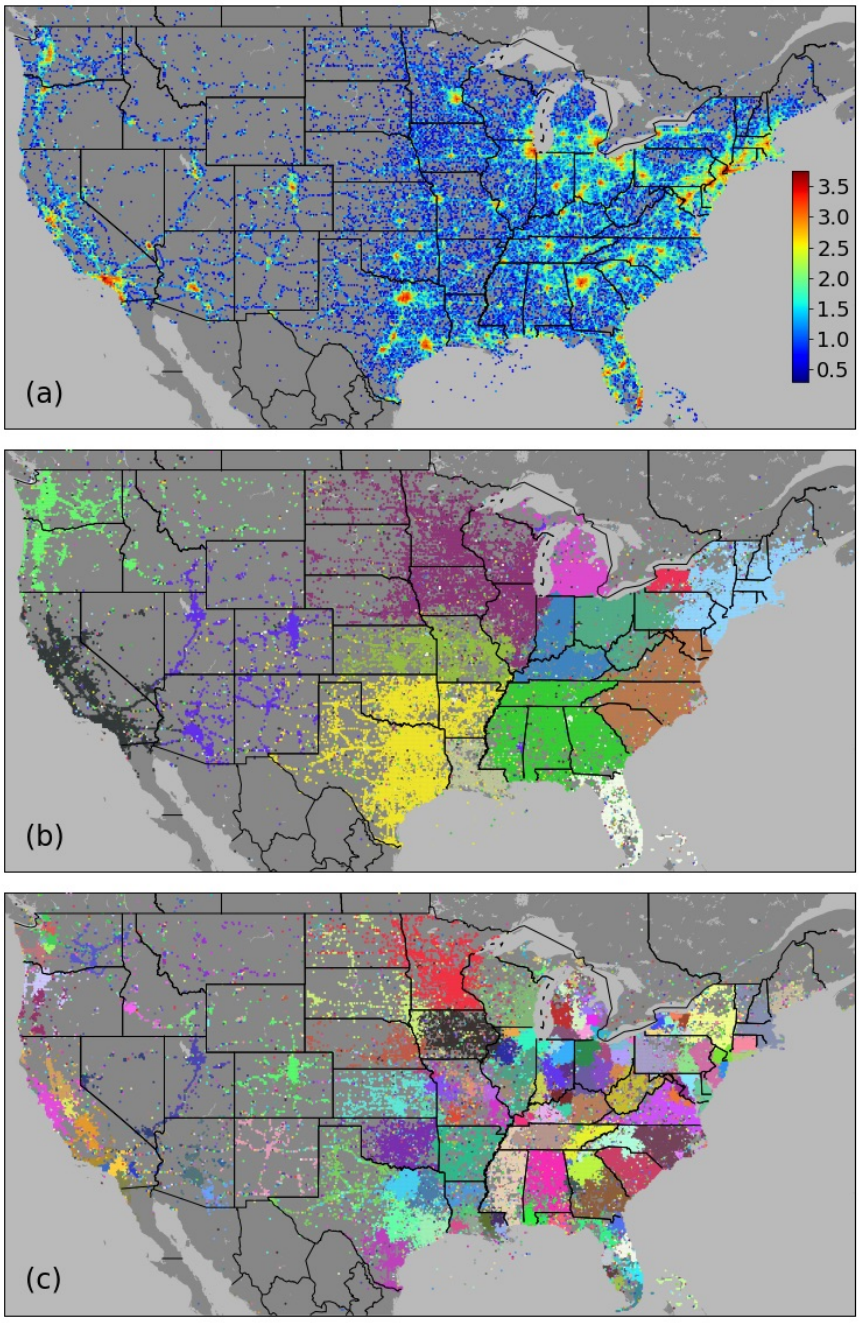


Figure 2: Structure and fragmentation patterns of the network associated with human communication. (a) Spatial degree centrality of the communication network. Colors indicate the amount of communication at each location, measured by the logarithm of the degree centrality of each node (scale inset). The communication network was used to generate communities using modularity optimization, with distinct colors indicating (b) 15 patches that can be visually associated to states or regions and (c) 168 smaller sub-communities within the communities of panel (b).

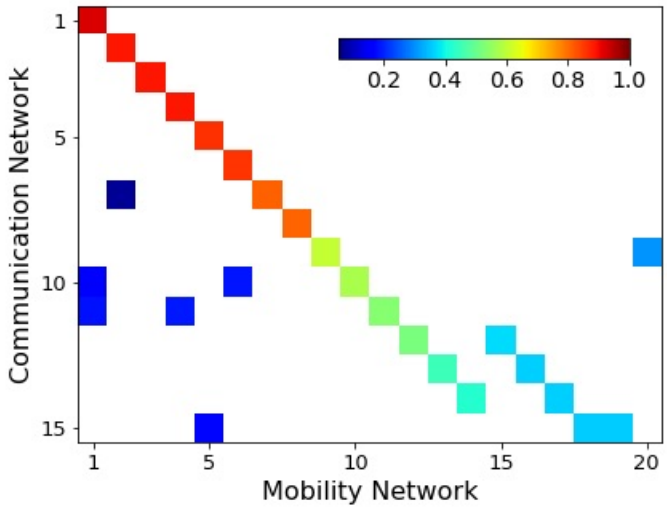


Figure 3: Similarity of communities in the communication and mobility networks. Matrix of the regional communities for the communication network (y-axis, $n = 15$) and mobility network (x-axis, $n = 20$), ordered by decreasing overlap between communities. Cell colors represent the number of nodes overlapping between the two networks in each community, normalized by the size of the communities per row (scale inset), with no overlap indicated in white.

community.

The communications network obtained from Twitter mentions is shown in Fig 2. Our modularity analysis on this network shows that it also has a spatial fragmented structure which is consistent with the mobility network. Thus, while the Internet and social media have drastically affected the dynamics of communications, the geographic structure of online communication remains fragmented and presents a similar structure to the one obtained from offline interactions. There are some differences. In contrast to the 20 modules in the mobility network, there are 15 modules that arise in the communication network. The borders of some communities in this figure are almost the same as those in the mobility network (Figure 1-b), such as the communities of Washington and its neighboring states, or the communities of Michigan and Florida. Ohio, western Pennsylvania and West Virginia are also still in the same patch. Some communities merge into a larger community in the communication network. For example, the New England area down to Washington, DC is now one single community. North and South Carolina and the patch including Virginia and Maryland are also combined together. This demonstrates that certain areas have a broader radius of online communication than physical travel. Finally, Figure 2-c represents the smaller communities within each community in Figure 2-b. These patches show areas connected to urban centers and are very similar to Figure 1-c. Some less populous states are now single communities such as Montana, Nebraska, Kansas, Oklahoma, Arkansas and New Mexico, while more densely populated areas are subdivided around urban centers.

In Figure 3, we quantitatively compare the modular structure of the mobility and communication networks by creating a matrix where we count the number of overlapping nodes of the

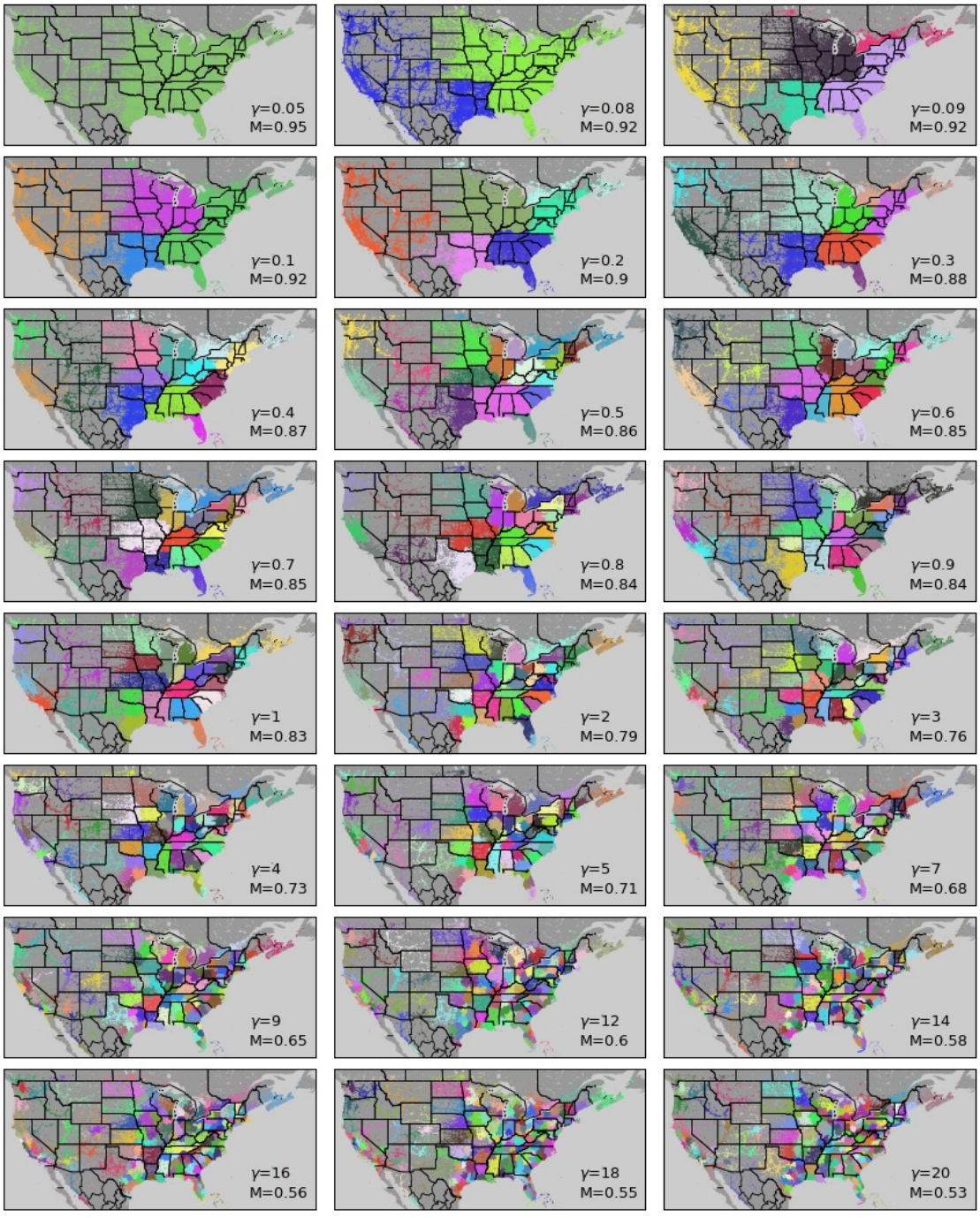


Figure 4: Multi-scale decomposition of the mobility network. Colors indicate geographical patches detected in the mobility network for values of the resolution parameter γ varied from $0.05 - 20$ (upper left to bottom right). M stands for modularity.

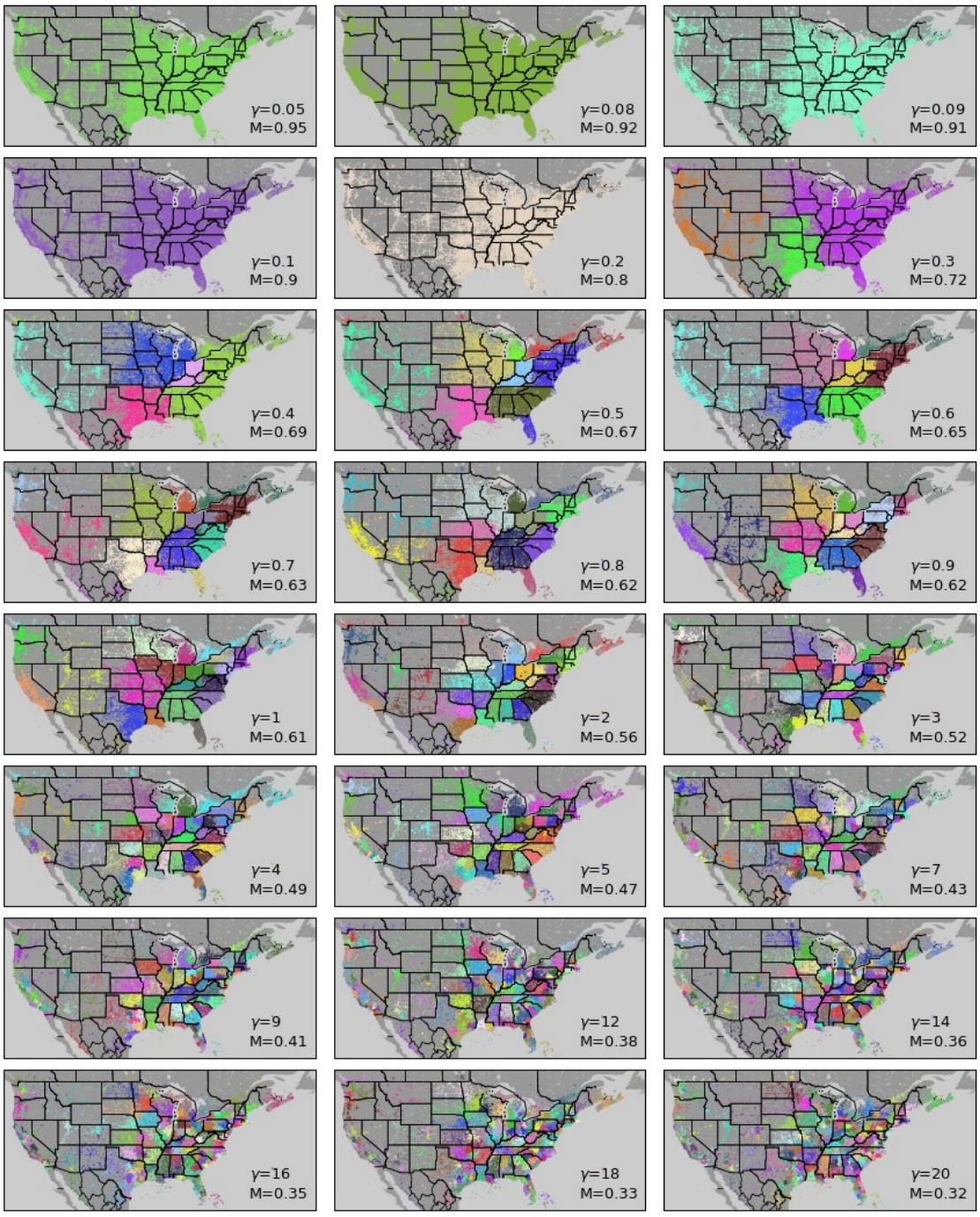


Figure 5: Multi-scale decomposition of the communication network. Colors indicate geographical patches detected in the communication network for values of the resolution parameter γ varied from $0.05 - 20$ (upper left to bottom right). M stands for modularity.

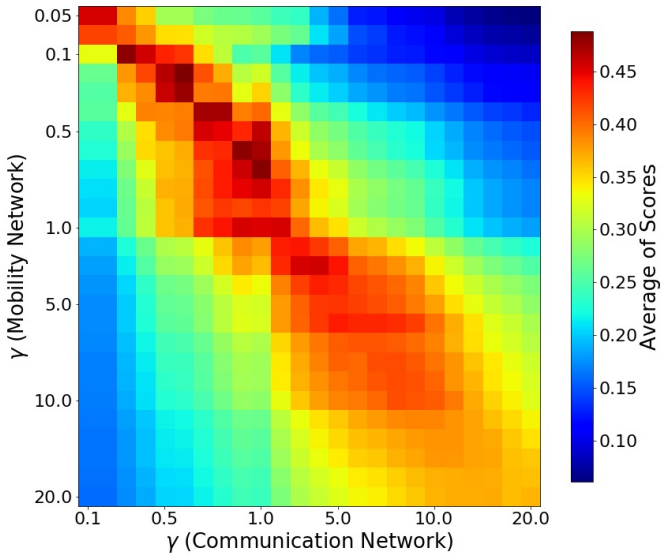


Figure 6: Similarity between the mobility and communication networks across multiple scales. Similarity is measured by the average of the Purity, Adjusted Rank, and Fowlkes-Mallows Indexes (color scale shown). Scale is defined by the different values of the resolution parameter γ (horizontal and vertical axes).

communities. Rows have been normalized by the size of each community in the communication network. According to the figure, some communities from the communication network are almost identical in the mobility network and therefore show a high overlap (red). Others are similar but not identical. A few communities from the mobility network are merged into communities in the communication network (green and light blue). Despite the observed differences in the networks representing two fundamentally different types of interaction, the modular structure is remarkably consistent, revealing that there is a strong coupling between the way people travel in physical space and communicate with each other online.

In order to further understand similarities between the mobility and communication networks, we perform a multi-scale analysis of their community structure using a generalized modularity optimization algorithm that introduces a resolution parameter, γ [36]. Smaller values of γ identify progressively larger communities, and vice-versa. The multiscale analysis of the mobility and communication networks are shown in Figures 4 and 5 respectively. Partitions range from a single large module of the entire US (top panels in Figure 4), down to urban scale partitions (bottom panels). Some states like Pennsylvania are split into multiple communities early in the process ($\gamma \approx 0.3$), while other states like Texas first emerge as single communities ($\gamma \approx 0.6$) and internally fragment later in the process ($\gamma \approx 0.1$). These differences are directly associated with the internal structure of social ties and their geographical breakpoints.

We compare the partitions in both networks for different values of γ by using three measures of cluster similarity: Purity [37], Adjusted Rand Index [38] and Fowlkes–Mallows Index [39]. These measures evaluate the overlap of partitions, with values ranging between 0 (no intersection)

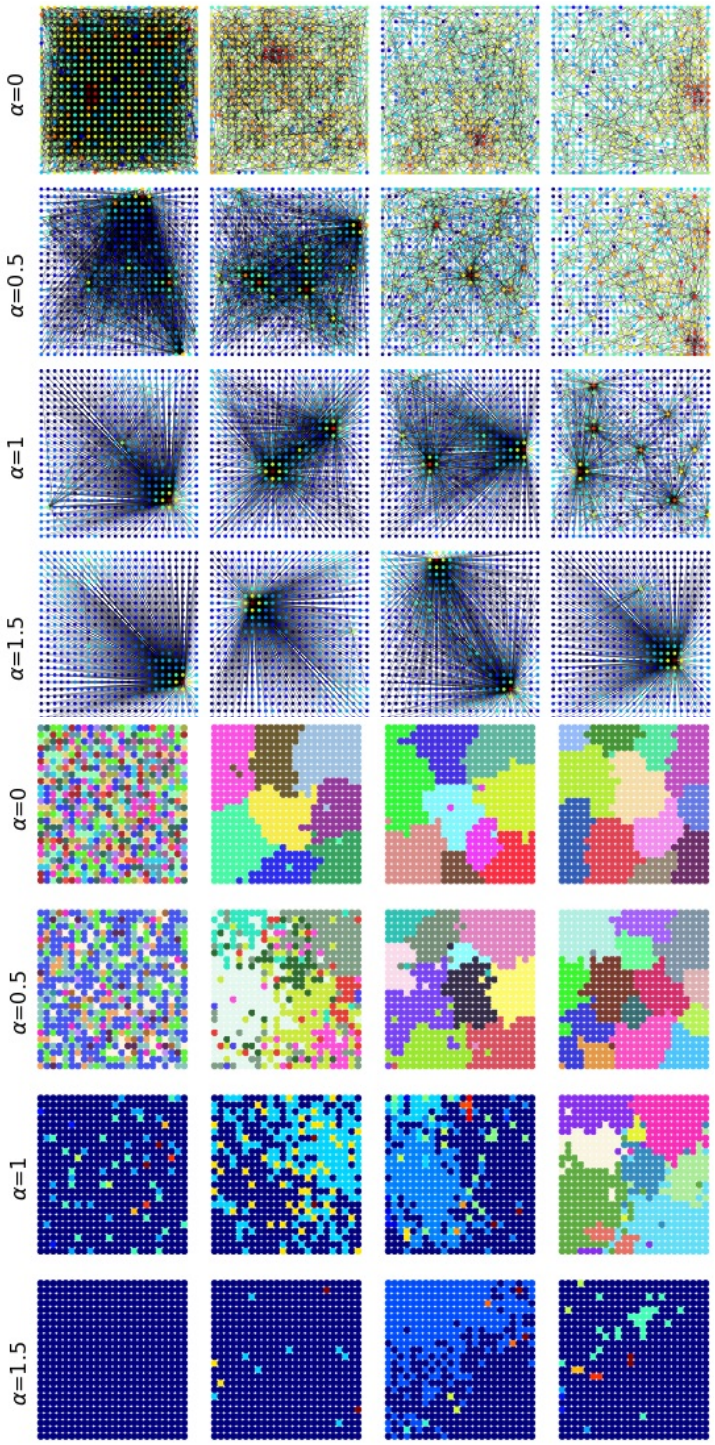
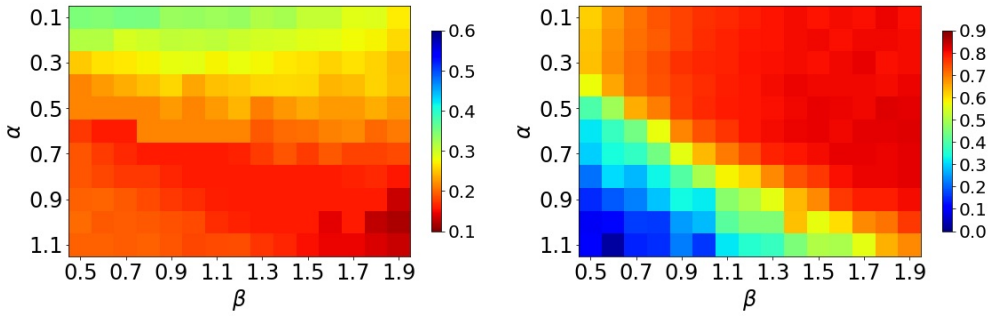


Figure 7: Spatial degree distribution and modular structure for model simulations with different parameter values α and β and a fixed value of $\nu = 0.1$ (see text for parameter definitions). Top panels show the spatial degree distribution (from weakly connected in blue to highly connected in red). Bottom panels show the modules of each graph—each color identifies a single community.



(a) K-S test (b) Modularity

Figure 8: Kolmogorov-Smirnov score and modularity for simulations with varying model parameter, α and β, and fixed resolution parameter $\nu = 0.1$. (a) Colors indicate the Kolmogorov-Smirnov (K-S) score, with lower scores (red) indicating similarity between the degree distributions of the model and the mobility network. (b) Colors indicate network modularity. Modularity is highest at around 0.8 (dark red), similar to the actual modularity for the mobility network (0.83).

and 1 (perfect match). Figure 6 shows a matrix whose rows and columns represent the partitions of the mobility and communication networks at different values of resolution (γ-communication and γ-mobility) and whose elements show the average of the three measures of similarity. The highest similarity between the two networks occurs at similar values of resolution (red diagonal), showing that the relative structure of these networks is consistent across scales (see Figure S3 for further analysis).

The consistency between the mobility and communication networks reveals that social spaces are not limited to the physical space. Instead offline interactions seem to condition the structure of online communications. Moreover, the hierarchical multiscale structure of these networks reveals that smaller regions of cohesive social ties, interactions and association belong to progressively larger ones. It may be expected for small communities from the same large community show progressively more similar behaviors and identity than communities of similar size from separate larger regions.

We constructed a network growth model that combines aspects of network dynamics and human mobility in order to show the emergence of spatial fragmentation. Our model combines preferential attachment [40], geographical distance gravity [41] and spatial growth. We begin with a lattice representing geographical locations, and grow connections among them simulating the way people travel. The probability of creating an edge between locations i and j in each time step is:

$$P_{ij} \sim < k_{nn} >_i^{\nu} \frac{k_j^{\alpha}}{d_{ij}^{\beta}} \tag{1}$$

where i represents the origin of the interaction, j indicates the destination, $< k_{nn} >_i$ indicates i's nearest neighbors' average degree, k_j represents j's degree and d_{ij} represents the distance between i and j. The exponents α, β and ν respectively control the effects of the preferential attachment mechanism, geographical distance gravity and spatial growth. The model reproduces the growth of cities (ν), their degree of attractiveness (α) and the linkage between urban center

and surrounding areas, including neighboring cities (β).

Figure 7 shows the results of model simulations in terms of the spatial degree distribution (top panels) as well as modular structure (bottom panels) for different values of α (rows) and β (columns) and a fixed value of $\nu = 0.1$. If we do not include the effects of either preferential attachment ($\alpha = 0$) or gravity ($\beta = 0$), the destination of edges are independently distributed among all nodes and the resulting communities have no spatial pattern. If $\alpha > \beta$, then a few hubs and one or two communities arise without significant geographic effects. Spatial fragmentation arises when the gravity mechanism is stronger than the preferential attachment ($\beta > \alpha$), either without hubs ($\alpha = 0$) or with hubs ($\alpha > 0$). Increasing ν leads to more localized high-activity areas (cities), but this also destroys localized patches, leading to lower values of modularity (see the Supplement for results exploring variation of the spatial growth mechanism in Figure S4).

We validated the model results against Twitter data by first testing whether the degree distributions from both sources are drawn from the same distribution and second comparing the modularity values. For each set of parameters, we created 20 model realizations and analyzed their statistical behavior. We applied the Kolmogorov-Smirnov statistical test (K-S) to compare the average degree distribution from the model realizations to the mobility network, and similarly to the communication network. Figure 8-a shows the values of the test results for different values of α and β (rows and columns of the matrix) and $\nu = 0.1$. Lower K-S values (red) indicate more similarity and higher K-S values (blue) indicate less similarity. The average modularity values for the simulations in Fig. 8-a are shown in Fig. 8-b, ranging from 0 (no modular structure) to 1 (high modular structure). We find that $\alpha = 0.9$ and $\beta = 1.5$ and $\nu = 0.1$ give a good fit between simulations and observed data (see Figures S5 and S6).

Understanding the structure and dynamics of groups is an essential aspect of understanding social interactions generally. The functioning of human societies arises not only from the activities of individuals but also from their interaction and integration by means of social ties. We analyzed the structure of social ties in the US using Twitter data and found multiscale, self-organized fragments that span from urban up to national scales for two types of interactions, mobility and communication. Our results show that the structures emerging from both types of interaction are highly consistent, revealing that social ties couple the integration and separation of groups in both physical and virtual spaces.

We constructed a model of network growth that is consistent with the emergence of the observed patterns. The model shows that spatial fragmentation may result from just short-distance travel, but this mechanism alone does not explain the emergence of highly connected places such as cities. Cities emerge from preferential attachment and spatial growth mechanisms, which increase heterogeneity in the degree distribution but may destroy spatial patches. Therefore, a combination of the three mechanisms replicate both heterogeneous degree distributions and spatial fragmentation.

The formation of groups and their interactions are intimately related to the formation of individual identity through self-identification and adoption of group norms and narratives. Thus, while individual identities are highly complex and unique, there are shared patterns among members of self-associating groups. These common patterns define the group identity, which may involve linguistic, cultural, economic, opinion or interest differences from other groups. The geographical

group structure can expose ethnic separation, class structure, and other reasons for social fragmentation.

Acknowledgements. We thank Rachel A. Rigg and Matthew Hardcastle for proofreading the manuscript.

References

[1] A.J. Morales, V. Vavilala, R.M. Benito, and Y. Bar-Yam. Global patterns of synchronization in human communications. *J. R. Soc. Interface*, 14(128):20161048, 2017.

[2] G. Krings, F. Calabrese, C. Ratti, and V. Blondel. Urban gravity: a model for inter-city telecommunication flows. *J. Stat. Mech.*, page L07003, 2009.

[3] G. D. Nelson and A. Rae. An economic geography of the United States: From commutes to megaregions. *PLoS One*, 11(11):1–23, 2016.

[4] A. Herdağdelen, W. Zuo, A. Gard-Murray, and Y. Bar-Yam. An exploration of social identity: The geography and politics of news-sharing communities in Twitter. *Complexity*, 19(2):10–20, 2013.

[5] M. Granovetter. The strength of weak ties: A network theory revisited. *Sociological Theory*, 1:201–233, 1983.

[6] E. Bakshy, S. Messing, and L. A. Adamic. Exposure to ideologically diverse news and opinion on Facebook. *Science*, 348(6239):1130–1132, 2015.

[7] A.J. Morales, J. Borondo, JC. Losada, and RM. Benito. Measuring political polarization: Twitter shows the two sides of Venezuela. *Chaos*, 25:3, 2015.

[8] T. C. Schelling. Dynamic models of segregation. *J. Math. Sociol.*, 1(2):143, 1971.

[9] M. Lim, R. Metzler, and Y. Bar-Yam. Global pattern formation and ethnic/cultural violence. *Science*, 317:1540–1544, 2007.

[10] A. Varshney. *Ethnic Conflict and Civic Life: Hindus and Muslims in India*. Yale University Press, 2008.

[11] D. L. Horowitz. *Ethnic Groups in Conflict: With a new preface*. Berkeley: University of California Press, 2011.

[12] M. J. Esman. *Ethnic conflict in the western world*. London: Cornell University Press, 1979.

[13] C. Herrera-Yagüe, C. M. Schneider, T. Couronné, Z. Smoreda, R. M. Benito, P. J. Zufiria, and M. C. González. The anatomy of urban social networks and its implications in the searchability problem. *Scientific Reports*, 5:10265, 2015.

[14] P. J. Mucha, T. Richardson, K. Macon, M. A. Porter, and J-P. Onnela. Community structure in time-dependent, multiscale, and multiplex networks. *Science*, 328(5980):876–878, 2010.

[15] H. Mao, X. Shuai, Y-Y. Ahn, and J. Bollen. Mobile communications reveal the regional economy in Côte d'Ivoire. *Proc. of NetMob*, 2013.

[16] C. Thiemann, F. Theis, D. Grady, R. Brune, and D. Brockmann. The structure of borders in a small world. *PLoS ONE*, 5:e15422, 2010.

[17] M. Fujita, P. Krugman, and A. J. Venables. The spatial economy: cities, regions, and international trade. *The MIT Press*, 2001.

[18] R. Lambiotte, V. D. Blondel, C. de Kerchove, E. Huens, C. Prieur, Z. Smoreda, and P. Dooren. Geographical dispersal of mobile communication networks. *Physica A: Statistical Mechanics and its Applications*, 387:5317–5325, 2008.

[19] M. Granovetter. The strength of weak ties. *American Journal of Sociology*, 78(6):1360–1380, 1973.

[20] R. L. Coser. *'Complexity of roles as a Seedbed of Individual Autonomy' in 'The Idea of Social Structure: Essays in Honor of Robert Merton'*. Harcourt Brace Jovanovich, 1975.

[21] I. Pool. *Comment on Mark Granovetter's 'The Strength of Weak Ties: A Network Theory Revisited.'*. Read at the 1980 meeting of the International Communications Association, Acapulco, Mexico, 1980.

[22] F. Noah. A test of the structural features of Granovetter's 'strength of weak ties' theory. *Social Networks*, 2:411–422, 1980.

[23] R. Breigern and P. Pattison. The joint role structure of two communities' elites. *Sociological Methods and Research*, 7(2):213–226, 1978.

[24] F. Calabrese, D. Dahlem, A. Gerber, D.D. Paul, X. Chen, J. Rowland, C. Rath, and C. Ratti. The connected States of America: Quantifying social radii of influence. *in Proc. of IEEE International Conference on Social Computing (SocialCom)*, 2011.

[25] T. Menezes and C. Roth. Natural scales in geographical patterns. *Scientific Reports*, 7:45823, 2017.

[26] A. Amini, K. Kung, C. Kang, S. Sobolevsky, and C. Ratti. The impact of social segregation on human mobility in developing and urbanized regions. *EPJ Data Sci.*, 3(6), 2014.

[27] F. Xiang, L. Tu, B. Huang, and X. Yin. Region partition using user mobility patterns based on topic model. *16th International Conference on Computational Science and Engineering*, 2014.

[28] C. Ratti, S. Sobolevsky, F. Calabrese, C. Andris, J. Reades, M. Martino, R. Claxton, and S. H. Strogatz. Redrawing the map of Great Britain from a network of human interactions. *PLoS One*, 5(12):e14248, 2010.

[29] S. Sobolevsky, M. Szell, R. Campari, T. Couronné, Z. Smoreda, and C. Ratti. Delineating geographical regions with networks of human interactions in an extensive set of countries. *PloS One*, 8:e81707, 2013.

[30] V. Blondel, G. Krings G, and I. Thomas. Regions and borders of mobile telephony in Belgium and in the brussels metropolitan zone. *Brussels Studies*, 42, 2010.

[31] S. Šćepanović, I. Mishkovski, P. Hui, J. K. Nurminen, and A. Ylä-Jääski. Mobile phone call data as a regional socio-economic proxy indicator. *PloS One*, 10:e0124160, 2015.

[32] D. Lazer, A. Pentland, L. Adamic, S. Aral, A-L. Barabási, D. Brewer, N. Christakis, N. Contractor, J. Fowler, M. Gutmann, T. Jebara, G. King, M. Macy, D. Roy, and M. Van Alstyne. Computational social science. *Science*, 323(5915):721–723, 2018.

[33] P. Deville, C. Song, N. Eagle, V. D. Blondel, A-L. Barabási, and D. Wang. Scaling identity connects human mobility and social interactions. *Proc. Natl. Acad. Sci. USA*, 113:7047–7052, 2016.

[34] Y. Bar-Yam. From big data to important information. *Complexity*, 21(S2):73–98, 2016.

[35] V. D. Blondel, J-L Guillaume, R. Lambiotte, and E. Lefebvre. Fast unfolding of communities in large networks. *J. Stat. Mech.*, page 10008, 2008.

[36] J. Reichardt and S. Bornholdt. Statistical mechanics of community detection. *Phys. Rev. E*, 74:016110, 2006.

[37] J. Artiles, J. Gonzalo, and S. Sekine. The Semeval-2007 Weps evaluation: Establishing a benchmark for the web people search task. *Workshop on Semantic Evaluation*, 2007.

[38] L. Hubert and P. Arabie. Comparing partitions. *Journal of Classification*, 2:193–218, 1985.

[39] E.B. Fowlkes and C.L. Mallows. A method for comparing two hierarchical clusterings. *Journal of the American statistical association*, 78(383):553–569, 1983.

[40] A-L. Barabási and R. Albert. Emergence of scaling in random networks. *Science*, 289:509–512, 1999.

[41] F. Simini, M. C. González, A. Maritan, and A-L. Barabási. A universal model for mobility and migration patterns. *Nature*, 484:96–100, 2012.

Supplementary Material

S1 Generalized modularity optimization method

We analyze social fragmentation by means of applying modularity optimization [35] to the mobility and communication networks obtained from Twitter data. Modularity is a scalar value $0 < M < 1$ that quantifies how distant the amount of edges inside a community is from that of a random distribution. The method iteratively finds a network partition that maximizes M.

To study communities at multiple scales, we use a generalized version of modularity [36] that includes a resolution parameter γ. In the conventional modularity equation, $\gamma = 1$, and the same weight is given to observed links and expected ones from a randomized network. In the generalized form, $\gamma < 1$ gives more weight to the observed links, which generates larger communities, while $\gamma > 1$ puts more weight on the randomized term and generates smaller communities. Because it is a method with multiple maxima, we chose partitions that are robust to multiple runs of the algorithm.

S2 Contribution of states to communities in the mobility and communication networks

To further quantify how state borders contribute to communities, we counted the number of nodes for each community in each state, and normalized to the size of the largest community. Figure S1 shows the contribution of states to communities for both the communication and mobility networks with resolution $\gamma = 1$. For many of the states, the borders of states match their communities, while other states have administrative borders that deviate from the ways people travel and communicate.

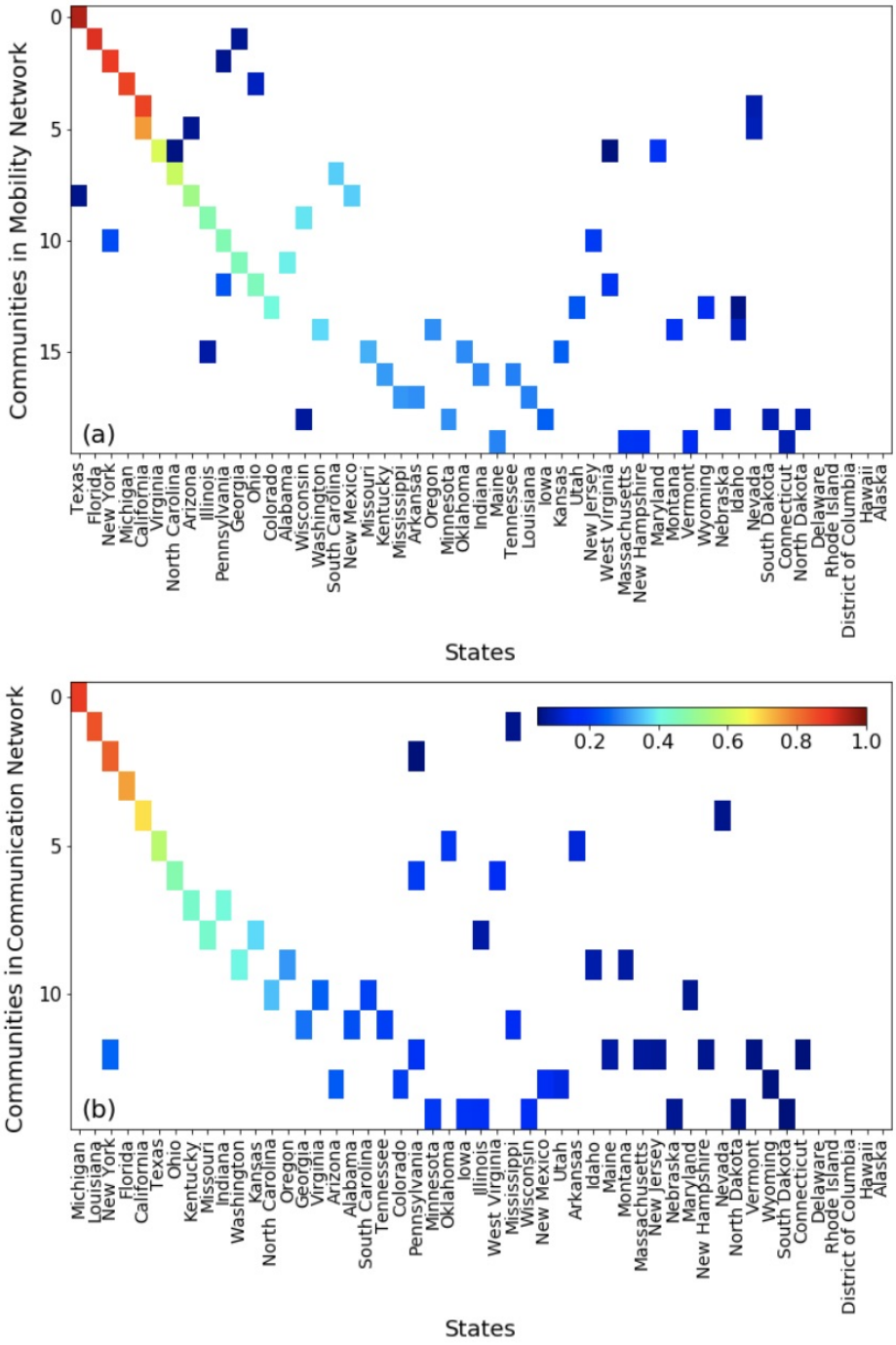


Figure S1: Contribution of states to communities. Graphs depict the number of communities detected (y-axis) in each state (x-axis) for (a) mobility and (b) communication networks generated with $\gamma = 1$. Colors in red indicate more overlap between state and community boundaries.

S3 Structural similarities of the mobility and communication networks

We compare structural properties of the mobility and communication networks by means of centrality measures, edge weights and multiscale structure. The structural properties of both networks are consistent with each other, showing an interplay between how people explore the physical space and communicate on Twitter.

In Figure S2-a we show a scatter plot of degree centrality for each location in both networks colored by their eigenvector centrality in logarithmic scale. While most locations are poorly connected, a few of them have an extremely high degree, corresponding to densely populated areas in large cities. Locations with a higher degree centrality in both communication and mobility networks also have high eigenvector centrality, which means that these locations are central relative to where information flows.

Next, we compare edges weight and length for both networks in Figure S2-b. The edge length is estimated as the geographical distance between the locations' centroids. Edges that have high weights also have small lengths, reflecting daily, short distance travels seen in cities and localized communication.

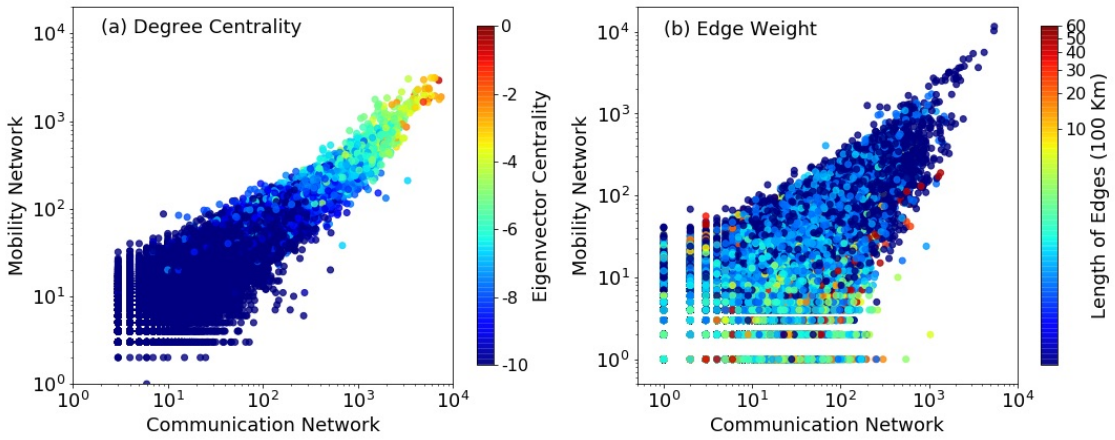
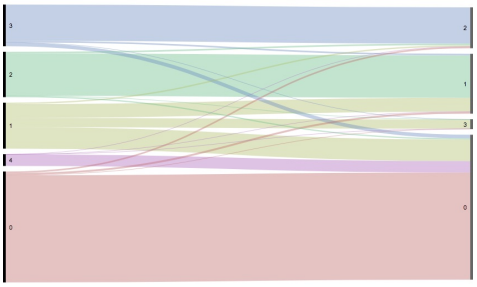


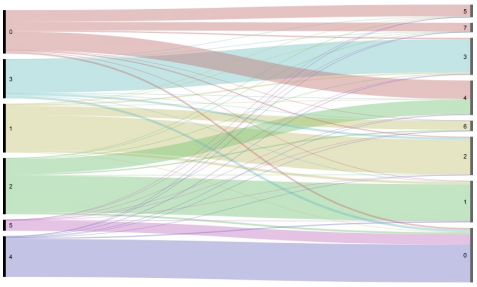
Figure S2: Similarity of network structure between the communication and mobility networks. (a) Scatter plot of degree centrality for each location in the mobility (y-axis) and communication (x-axis) network, colored by the corresponding eigenvector centrality. (b) Scatter plot of the edge weights for each location in the mobility (y-axis) and communication (x-axis) network, colored by the edge length or distance between nodes. Axes and colorbars are in logarithmic scale. Scales are shown on figure.

Finally, we compare the multiscale structure of the networks fragmentation. Fragmentation can be seen at multiple scales using the generalized modularity optimization method. This method seeks partitions at various scales by considering the resolution parameter (γ).

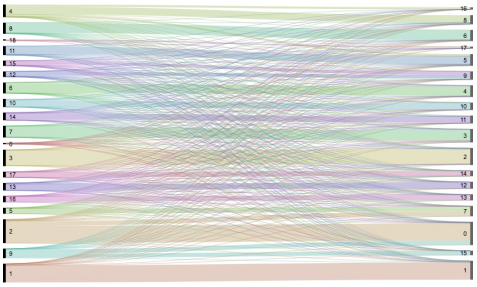
We compare the two networks at roughly similar partition sizes (mobility and communication respectively at γ=0.1 and 0.3, γ=0.2 and 0.6, γ=0.6 and 0.9, or γ=0.7 and 1.0). These pairs of γ



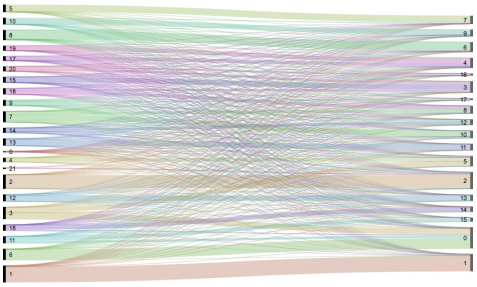
Mobility γ=0.1 - Communication γ=0.3



Mobility γ=0.2 - Communication γ=0.6



Mobility γ=0.6 - Communication γ=0.9



Mobility γ=0.7 - Communication γ=1.0

Figure S3: Comparison of the mobility and communication networks at γ values with highest similarity. At each panel, an alluvial diagram maps similarities of detected communities in the mobility (left axis) to the communication (right axis) networks. Values of γ are chosen from the four darkest red cells in Fig. 6, which yield the highest similarity between the two networks.

values represent partitions in both networks with the highest similarities (see Figure 6 in the main text). In Figure S3, we show the partitions comparison using an alluvial diagram for each set of γ values (panels). The alluvial diagrams map the corresponding number of nodes at each module of the mobility network (left axis) onto each module of the communication network (right axis).

S4 Additional results from varying the model parameters

In Figure S4, we show the effect of changing the spatial growth term ν in the network model. This term gives preference to locations in which the average degree of nearest neighbors is higher. We set the α and β exponents at 0.5 and 1.5, respectively, which generates geographical fragmentation patterns even with $\nu = 0$ (left panels). As shown in figure S4-(a), increasing the ν exponent (from left to right panels) concentrates the connections around hotspots, recreating the growth of cities. However, Figure S4-(b) and (c) show that while small values of ν increase the number of communities and their modularity, large values of ν lead to less cohesive borders between communities and a decrease in modularity.

Next, we investigated how changing all three parameters in the model causes deviations between the model and the real data. We used the Kolmogorov-Smirnov (K-S) test, a measure of similarity between two distributions, to determine the similarity of the network degree distributions, which are a measure of network connectivity. Degree distributions for the model were calculated by averaging over 20 realizations for each set of parameters. Figure S5 represents the similarity of degree distributions of locations for the model versus the empirical mobility network. Moving from the top left panel to the bottom right panel, we see that the degree distributions for the model and the mobility network data deviate from each other as the ν exponent increases.

We also determined how changing the model parameters affects the modularity of the network. Figure S6 shows the average modularity from 20 realizations for different values of the three exponents. Overall, increasing the strength of the preferential attachment process (controlled by α) and the spatial growth process (controlled by ν) destroys geographical patches and reduces modularity. Conversely, the human mobility gravity process (controlled by β) is the only exponent that increases these fragmented geographical patches. Note that modularity is 0.83 for the mobility network from the Twitter data.

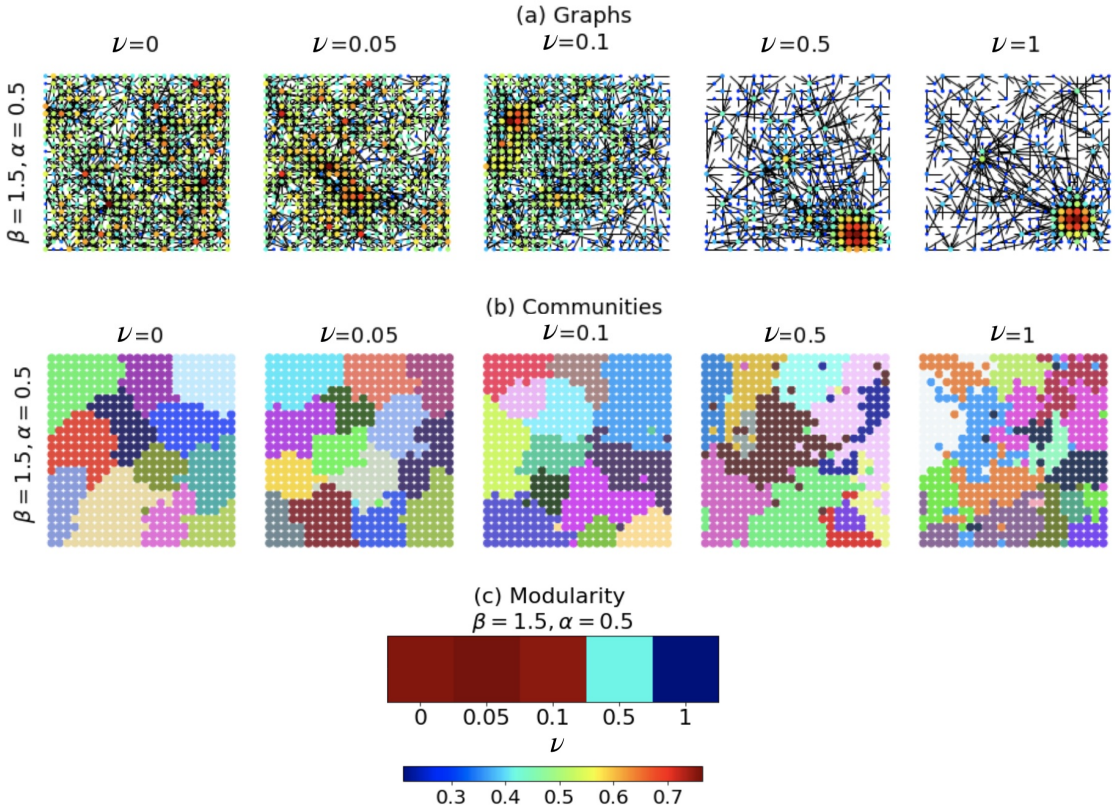


Figure S4: Degree centrality, fragmentation, and modularity for simulations with varied values of ν (see section S4). The resolution parameter ν is varied from $0, 0.05, 0.1, 0.5, 1$ (left to right), while $\alpha = 0.5$, $\beta = 1.5$, and the size of the lattice is 576 locations. (a) Spatial degree centrality. Nodes are colored by their degree centrality (from blue to red) and edges are plotted in black. (b) Spatial patches, shown in varying colors. (c) Modularity as a function of ν, indicated by color (scale below figure).

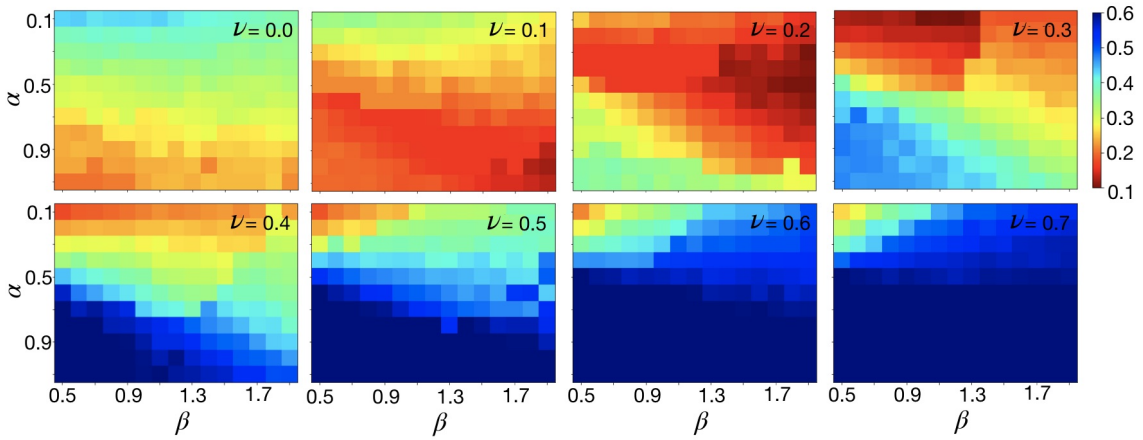


Figure S5: Similarity of degree distributions between simulated and real mobility networks. Matrices show different values of the parameters α (y-axis), β (x-axis) and ν (panels, upper left to lower right). Kolmogorov-Smirnov (K-S) scores are depicted with colors (scale on right), with the lowest K-S values in red, indicating that the degree distributions are similar between simulated and real mobility networks.

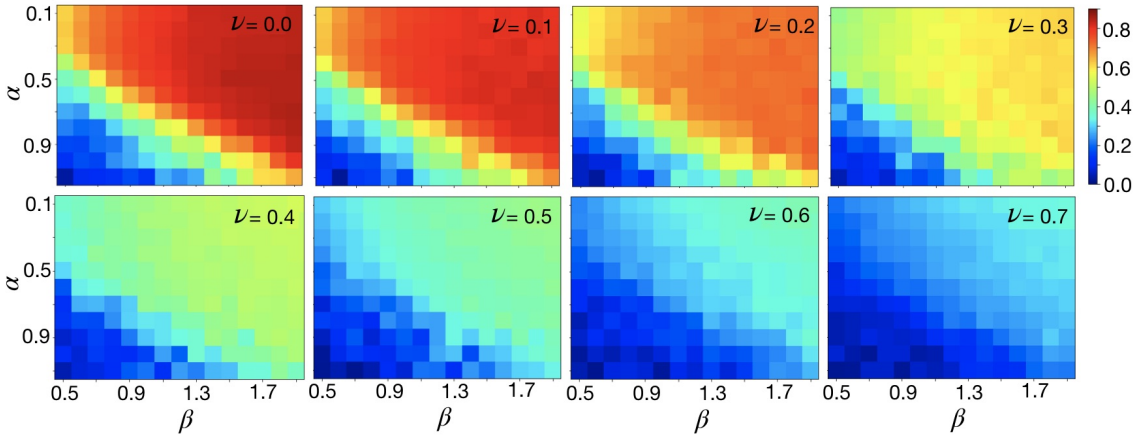


Figure S6: Modularity of detected communities in simulations with varying model parameters. Matrices show different values of the parameters α (y-axis), β (x-axis) and ν (panels, upper left to lower right). Modularity values are depicted with colors (scale on right), with the highest values in red at ~ 0.9. Note that modularity for the mobility network is 0.83.